CHEMNITZ UNIVERSITY OF TECHNOLOGY

Faculty of Natural Sciences

Department of Physics

Computational Physics

Bakkalaureusarbeit Computational Science

Optimal Move Class For Simulated Annealing With Underlying Optimal Schedule

Ines Hartwig

Chemnitz, July 26, 2005

Hartwig, Ines
ines.hartwig(at)s2002.tu-chemnitz.de
Optimal Move Class For Simulated Annealing With Underlying Optimal Schedule
Bakkalaureusarbeit
Department of Physics,Chemnitz University of Technology, March 2009

Acknowledgement

My sincere thanks go to all the people who contributed to this thesis each in their own way. In particular, I am grateful for the help of Frank Heilmann and the patience of Prof. Dr. K. H. Hoffmann.

Abstract

Die vorliegende Arbeit befasst sich mit dem Versuch der Optimierung von Simulated Annealing. Genauer gesagt, werden Simulationsergebnisse für einfache Spinglassysteme in Abhängigkeit von verschiedenen Nachbarschaftsmodellen berechnet – jeweils unter Verwendung des optimalen Abkühlverlaufs. Ziel ist es, eine Faustregel für die dynamische Anpassung der Nachbarschaftsbeziehung während einer Annealing-Simulation zu finden.

The thesis at hand presents an attempt to optimize simulated annealing. In particular, annealing results are computed based on different move class definitions for Ising spin systems while simultaneously applying an existing algorithm to determine the optimal temperature schedule for each case. The aim is to find a rule of thumb for dynamic adjustment of the move class during an annealing run.

Contents

List of Figures

List of Tables

1 Introduction

Simulated annealing is widely used in stochastic optimization of complex systems featuring solution spaces with many local optima; problems too complex to be solved analytically or by means of conventional, deterministic, numerical methods. Apart from the fact that very specified algorithms perform better for their one specific problem, simulated annealing has the advantage that it is a *general* method, i. e. it can be applied to manifold problems in various fields. Annealing simulations have proved useful for such well-known issues as the travelling salesman problem, chip layer design in microelectronics, optimization of learning procedures for neural networks and of course, as the principles originate from physics, finding ground or energetically low-lying states of physical systems.
The main concept of the algorithm is that of a random walk through the solution space directed by:

- a move class, assigning each solution a set of neighboring solutions the random walker may try to move to.
- a rule defining an acceptance probability for each move depending on a control parameter,
- a schedule of how the control parameter (temperature) is to be lowered during the simulation,

This thesis will investigate the influence of modifications to the move class on annealing results for the example of Ising spin systems. In order to achieve comparability, it is necessary to base the simulations each on the same acceptance rule and on their optimal temperature schedule, so as to neglect any effects due to "unfit cooling." An existing algorithm to determine these schedules uses the description of the random walk by the master equation. It will be slightly modified here – in the sense that the primordial transition probability matrix has to be adjusted to the respective move class order.

The so collected results will be interpreted in order to find a way of run-time adjustment of the move class. The right choice of an annealing schedule and a clever move class definition can be of great impact on how near the random walker is to the

global optimum in the end. In that way, the adjustment of these two things may be viewed as an attempt to optimize optimization. Hopefully, the results obtained can give relevant hints how to improve performance of future annealing simulations.

2 The Principles Of Simulated Annealing

The principles of simulated annealing are derived from equilibrium statistical mechanics, which therefore will be the topic of the first part of this overview. The second part contains the description of the method itself.

2.1 Equilibrium Statistical Mechanics

An ergodic system featuring discrete states α with corresponding energies E_α is ruled by the Boltzmann-Gibbs statistics for thermal equilibrium. For a system at any temperature T, with k_B denoting Boltzmann's constant, the partition function Z:

$$Z(T) = \sum_\alpha e^{-\frac{E_\alpha}{k_B T}},$$

can be used to give virtually *any* equilibrium property of the system [1]. [1] A popular example is the inner energy of the system– the statistical mean energy. With the probability of a state

$$p_\alpha(T) = \frac{e^{-\frac{E_\alpha}{k_B T}}}{Z}, \quad \sum_\alpha p_\alpha = 1,$$

the system's inner energy can be calculated as follows:

$$\overline{E}(T) = \sum_\alpha p_\alpha E_\alpha = -\frac{\partial}{\partial \frac{1}{k_B T}} \ln Z.$$

Generally, the expectation value of any arbitrary observable A is:

$$\overline{A} = \sum_\alpha p_\alpha A_\alpha.$$

For $T \to +\infty$, $p_\alpha \to 1/(\#\text{ states})$, i. e. the states are uniformly populated in the stationary (equilibrium) case. Contrarily, for $T \to 0^+$, $p_\alpha \to \delta_{\alpha,\,\mathrm{GS}}$, where the index

[1] Please note that $\sum_\alpha$ sums up over all linear independent *eigenstates* α and not the energy levels E_α which can be degenerated. Degenerated systems will not be investigated here.

"GS" signifies that of the system's ground state, i. e. the state with minimal potential energy. This temperature dependant behavior is understood as a result of the rivalry between the increasing thermal movements of the system's particles, gaining kinetic energy proportional to $k_B T$, and the stability criteria of the potential.

Accordingly, in experimental physics, information about a system's ground state may be collected by cooling it down from high temperature to near the absolute zero. Practically, the temperature has to be lowered slowly and carefully in order to imitate a quasi-stationary process allowing the system to equilibrate at each temperature and thus to prevent it from getting stuck in a metastable configuration with a local energy minimum.

The principle of slow temperature reduction so as to find global minima of complex configuration spaces is the brilliant idea behind the computational method of simulated annealing.

2.2 The Simulated Annealing Method

Simulated annealing is capable of finding near global minimum solutions to problems exhibiting a complex, multidimensional solution space with many local minima.

2.2.1 The Conception Of The Random Walk

As stochastic optimization algorithm, the annealing process itself may be imagined as a random walk exploring the solution space subject to certain restrictions so as to assure that the walker's final position is near to the global optimum concerning solution values:

(1) Starting with an initial solution β, e.g. generated randomly following a uniform distribution,

(2) a neighboring solution α is determined where the random walker shall *attempt* to jump to.
(2) will hence be referred to as a *choice.*

(3) The acceptance rule depending on the difference in a predefined cost function, $\Delta E = E_\beta - E_\alpha$, and a control parameter, T, is applied to decide whether the random walker steps to this neighbor, accepting it as the next β or stays at its current position.
(2)–(3) will hence be called a *move.*

(4) Ideally, moves are done until a stationary distribution is reached, which means the system has equilibrated at T. Practically, the progress can be monitored by simultaneously computing the mean cost function value's pattern. In the stationary case, $\overline{E}$ should stay level.

(5) Having reached equilibrium, T is decremented according to the defined schedule and the acceptance process is again repeated until the system equilibrates.
Steps (2)–(5) will be called an *annealing or time step.*

(6) T is lowered until the maximum number of time steps allowed is reached and the current solution is taken as final solution to the problem.

Due to time limitation in practice, a couple of variants using faster temperature reduction exist, not allowing the system to reach the stationary distribution. These are called rapid cooling, or – in the special case of $T = 0$ for all time steps – quenching.

2.2.2 The Governing Specifications

Considering (1)-(6) above, it is clear that several decisions have to be made in advance. Here are the main four things affecting the manner of the random walk [2]:

a) The cost function E defined on the solution space: Each solution α has a corresponding value E_α. In physics, E is usually the inner energy.

b) The neighborhood relation N or move class, i. e. the rule how to generate a neighboring solution $\alpha \in N(\beta)$ to the current solution β. Note that this relation is symmetric:
$\alpha \in N(\beta) \Leftrightarrow \beta \in N(\alpha)$. Thus, the states in state space are interconnected by an undirected graph structure.

c) The acceptance rule to decide whether α is adapted as next solution, β. Else, the current solution remains, succeeding itself.

d) The annealing schedule,i. e. the way in which the control parameter T is to be altered chronologically.

Point a) above is defined by the objective of the problem itself. Adjusting b), c) and d) heuristically, manifold problems in a variety of fields have been subject matter of successful annealing simulations:
The idea originates from physics, where Metropolis *et. al.* invented it in 1953 [3], using

Boltzmann-Gibbs statistics to simulate thermal equilibrium behavior of physical systems and a modified Monte-Carlo method to compute the state sums $\sum_\alpha e^{-E_\alpha/k_BT} \ldots$ – which in the classical limit for N particles pass into continuous $6N$-dimensional integrals.
Since already moderate macroscopic physical systems consist of $N \gtrapprox 10^{23}$ particles, they used a model with several hundred particles only and applied periodic boundary conditions to imitate macroscopic behavior and to avoid surface effects.
Not before another thirty years had passed, Kirkpatrick *et. al.* [4] published an article exploiting its principles' analogies to those of non-physical optimization problems and showing its extraordinarily good performance for use in microelectronic circuit design. Table 2.1 summarizes these basic analogies arising.
Also in the 1980s, another generalization was released by Tsallis, who established a

Symbol	Non-physical Application	Physical Application
	solution space	state space
	optimal solution	ground state
α	solution	state
E	cost function	system energy
T	control parameter	temperature

Table 2.1: Analogies between physical and non-physical simulated annealing

new statistics for nonextensive physical systems [5] – including the Boltzmann-Gibbs version as limit case. The performance of Tsallis statistics for finding ground states of disordered systems has been the matter of various investigations since then, e. g. in [6].
In this thesis the original Metropolis procedure is used. That means it employs conventional statistics for the Metropolis acceptance rule:

$$P(\Delta E, T) = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ e^{-\Delta E/T} & \text{if } \Delta E > 0\,. \end{cases} \tag{2.1}$$

where, for computational convenience, $T \equiv k_BT$. The weighting of state transitions with $e^{-\Delta E/T}$ for finite temperature is called Boltzmannization [7] and accounts for the appropriate stationary distribution.
Interpreting rule (2.1) for the random walk procedure, the chosen neighbor α is accepted automatically, if the change from β to α produces a decrease in energy. Else, a random number ξ from a uniform distribution in $[0; 1]$ is compared to the value

of the Boltzmann factor $e^{-\Delta E/k_B T}$. If ξ is smaller than or equal to the Boltzmann factor, α is also accepted as new state β, else it is rejected and β remains the current state. In the special case $T \to \infty$, all attempted moves are accepted: $P(\Delta E, \infty) = 1$. Contrarily, for $T = 0$ only moves downward in energy are allowed: $P(\Delta E, 0) = 0$ generally for $\Delta E > 0$.

Checkpoints a) and c) above are given by the state energies of the investigated systems and the usage of the Metropolis rule, respectively. A next step on the way to improve performance of the random walk is to customize the temperature schedule d) to the given problem. This will be the topic of the next chapter, whereas this thesis on the whole is aimed at finding a rule for adjustment of the neighborhood definition b).

3 Determination Of Optimal Schedules

Note that this chapter renders almost exactly the same description as in [8].

3.1 Description Of Simulated Annealing Via Master Equation

Mirroring [8], where the stochastic process of the random walk is illustrated by the master equation

$$\underline{p}^{\,t+1} = \underline{\underline{\Gamma}}\left(T^{t+1}\right)\underline{p}^{\,t}, \tag{3.1}$$

this section is about an algorithm to determine the optimal temperature schedule for a given state space and restricted number of annealing steps; a derivative of optimal control theory.

In equation (3.1), $\underline{p}^{\,t}$ is the over-all probability distribution of states at time step t with $\sum_\beta p_\beta^{\,t} = 1$. Consequently,$p_\beta{}^t$ denotes the probability of being in state β at that time.
$\underline{\underline{\Gamma}}\left(T^{t+1}\right)$ is the transition probability matrix: Being in state β at time t, the probability of having jumped to state α at time $t+1$ equals $\Gamma_{\alpha\beta}\left(T^{t+1}\right)$. $\underline{\underline{\Gamma}}(T)$ itself is composed of two factors containing neighborhood relation, $\underline{\underline{\Pi}}$, and the Metropolis acceptance rule, $P(\Delta E, T)$:

$$\Gamma_{\alpha\beta}(T) = \Pi_{\alpha\beta} \times P(\Delta E, T) \text{ for } \alpha \neq \beta \tag{3.2}$$

Note, that the diagonal elements are: $\Gamma_{\alpha\alpha} = 1 - \sum_{\zeta\neq\alpha}\Gamma_{\zeta\alpha}$.

The matrix $\underline{\underline{\Pi}}$ has non-zero entries $\Pi_{\alpha\beta}$ only for states α in the neighborhood of β:

$$\Pi_{\alpha\beta} = \begin{cases} c\begin{cases} 0 & \text{if } \alpha \notin N(\beta), \\ g_\alpha & \text{if } \alpha \in N(\beta), \end{cases} \\ 1 - \sum_{\zeta\neq\alpha}\Pi_{\zeta\alpha} \quad \text{for } \alpha = \beta, \end{cases} \tag{3.3}$$

where $c = const$ sets the system's time scale[1]. For the non-degenerated systems considered here, the degeneracies $g_\alpha = 1\ \forall\ \alpha$ and $c = 1/|N(\beta)|$. $|N(\beta)|$ is the

[1] See section 5.3.1

number of state β's fundamental neighbors which is the same for all states. Since $\underline{\underline{\Pi}}$ – as well as $\underline{\underline{\Gamma}}$ – is a stochastic matrix, the column sums $\sum_\alpha \Pi_{\alpha\beta} = \sum_\alpha \Gamma_{\alpha\beta}$ have to be normalized to 1: Evidently, the random walker *must* proceed to *some* state! Another important fact is, that for $T \to +\infty$ *all* attempts are successful. This leads to

$$\underline{\underline{\Gamma}}\,(+\infty) = \underline{\underline{\Pi}},$$

and $\underline{\underline{\Pi}}$ turns out the probability matrix for the neighbor choice, as expected.

3.2 Schedule Determination

The actual algorithm was derived using the principles of variation:

> For given initial distribution $\underline{p}^{\,0}$ the value of an objective function f of the form
>
> $$f = \underline{E}^{\mathrm{T}}\,\underline{p}^{\,n}$$
>
> has to be minimized within a number n of annealing steps. The master equation is included as side condition with Lagrange multipliers $\underline{\Lambda}$ as follows:
>
> $$\begin{aligned} f &= \underline{E}^{\mathrm{T}}\underline{\mathrm{p}}^{\,n} + \sum_{t=0}^{n-1} \left(\underline{\Lambda}^{t+1}\right)^{\mathrm{T}} \left(\underline{\underline{\Gamma}}\left(T^{\,t+1}\right)\underline{\mathrm{p}}^{\,t} - \underline{\mathrm{p}}^{\,t+1}\right) \\ &= \left(\underline{E}^{\mathrm{T}} - \left(\underline{\Lambda}^{n}\right)^{\mathrm{T}}\right)\underline{p}^{\,n} - \sum_{t=1}^{n-1} \left(\underline{\Lambda}^{t}\right)^{\mathrm{T}}\underline{p}^{\,t} + \sum_{t=0}^{n-1} \left(\underline{\Lambda}^{t+1}\right)^{\mathrm{T}}\underline{\underline{\Gamma}}\left(T^{\,t+1}\right)\underline{\mathrm{p}}^{\,t} \end{aligned}$$
>
> The first variation is thence:
>
> $$\begin{aligned} \delta f = &\left(\underline{E}^{\mathrm{T}} - \left(\underline{\Lambda}^{n}\right)^{\mathrm{T}}\right)\delta\underline{p}^{\,n} + \sum_{t=1}^{n-1} \left(\left(\underline{\Lambda}^{t+1}\right)^{\mathrm{T}}\underline{\underline{\Gamma}}\left(T^{\,t+1}\right) - \left(\underline{\Lambda}^{t}\right)^{\mathrm{T}}\right)\delta\underline{p}^{\,t} \\ &+ \sum_{t=1}^{n-1} \frac{\partial\left(\underline{\Lambda}^{t+1}\right)^{\mathrm{T}}\underline{\underline{\Gamma}}\left(T^{\,t+1}\right)\underline{p}^{\,t}}{\partial T^{\,t+1}}\delta T^{\,t+1} \end{aligned}$$
>
> From the condition that $\delta f = 0$ in the minimum, one obtains that:
>
> $\underline{\Lambda}^{n} = \underline{E}$,
> $\underline{\Lambda}^{t} = \left(\underline{\underline{\Gamma}}\left(T^{\,t+1}\right)\right)^{\mathrm{T}}\underline{\Lambda}^{t+1}$ and
> $\underline{\Lambda}^{t+1}\underline{\underline{\Gamma}}\left(T^{\,t+1}\right)\underline{p}^{\,t}$ has to be in a minimum respective to $T^{\,t+1}$.

This gives rise to an iterative algorithm to determine the optimal temperature schedule, as rendered in [8]:

1. Compute $\underline{p}^{\,t+1,\ i=0} = \underline{\underline{\Gamma}}\left(T^{\,t+1,\ i=0}\right) \underline{p}^{\,t,\ i=0}$ for $t = 0, 1, \ldots,\ n-1$.
2. Compute $\underline{\Lambda}^{t-1,\ i} = \left(\underline{\underline{\Gamma}}\left(T^{\,t,\ i}\right)\right)^{\mathrm{T}} \underline{\Lambda}^{t,\ i}$ for $t = n, n-1, \ldots,\ 2$.
3. Compute $T^{\,t+1,\ i+1}$ such that $\left(\underline{\Lambda}^{t+1,\ i}\right)^{\mathrm{T}} \underline{\underline{\Gamma}}\left(T^{\,t+1,\ i+1}\right) \underline{p}^{\,t,\ i+1}$ has a minimum and determine $\underline{p}^{\,t+1,\ i+1} = \underline{\underline{\Gamma}}\left(T^{\,t+1,\ i+1}\right) \underline{p}^{\,t,\ i+1}$ for $t = 0, 1, \ldots,\ n-1$
4. Compare $f^{\,i+1}$ with the previous value $f^{\,i}$. If the difference is smaller than a chosen accuracy, then stop the iteration; otherwise increase i and go back to 2.

Note, that the initial distribution $\underline{p}^{\,0}$ and temperature schedule $\{T^{\,t,\ i=0}\}_{t=1,\ldots,\ n}$ have to be specified in advance for this to work, of course.

4 Ising Spin Systems

Spin systems based on the Ising model are experimented on here under modification of the move order. Note, that there is no immediate physical thought behind the specific model used, but the only aim was to construct state spaces with at least two local minima. Yet, in physics, the Ising model is widely used to understand the behavior of disordered systems, especially issues of paramagnetism.

4.1 The Ising Model

Generally, imagining a crystal lattice with L atoms numbered i and possessing spins s_i, the Hamiltonian looks like this [1]:

$$\widehat{H} = \sum_{i=1}^{L} \sum_{j=i+1}^{L} J'_{ij}\ \underline{\hat{S}}_i\ \underline{\hat{S}}_j - g\mu_B \underline{H} \sum_{i=1}^{L} \hat{S}_i{}^z. \tag{4.1}$$

$\underline{H}$ is the external magnetic field, assumed to be parallel to the z-axis, g is Landé's factor and μ_B Bohr's magneton. The J_{ij} rule the kind of interaction between the individual spins.
$\sum_{i=1}^{L} \sum_{j=i+1}^{L} \equiv \sum_{i<j}^{L}$ sums up the interaction over all pairs of spins.

The Ising model assumes only parallel or antiparallel alignment to the external field; that means $\underline{\hat{S}}_i\ \underline{\hat{S}}_j = \hat{S}_i{}^z \hat{S}_j{}^z$ and for electrons $< \hat{S}_i{}^z >= \pm\frac{1}{2} \equiv \frac{1}{2}s_i$. Consequently, equation (4.1) now gives the inner energy of the system in a state constituted by the whole of $\{s_k\}_{k=1,\ldots,L}$ as:

$$E_\alpha \equiv E(\{s_k\}) = -\sum_{i<j}^{L} J_{ij}\ s_i s_j - \mu H \sum_{i=1}^{L} s_i.$$

where $\mu = \frac{g\mu_B}{2}$. For computation, μH was set equal to 1 in terms of energy.
For an arrangement of L spins, the system features 2^L states, each spin being able to align either parallel, $s_i = 1$ or antiparallel $s_i = -1$ to the z-axis and $\underline{H}$.

4.2 Choosing Interactions

Often, the spins are depicted as assigned to sites (atoms) in a square lattice (crystal) of N sites periodically continued to limit surface effects. There exist manifold approaches of how to model the interaction between the spins for different physical systems, the most popular of which are probably those only taking into account a spin's 4 next neighbors (in a 2-dimensional square lattice), thus modelling short range interactions. However, the one here applied lets each spin s_i interact with each other s_j , $j \neq i$, allowing for the image of a chain of spins with length L. The J_{ij} are random numbers generated from a uniform distribution in $[-1; 1]$. They were chosen such that at least 2 local minima in energy exist. Practically, that means, the random interactions were generated and then in an exhaustive search – which for the small system sizes considered is absolutely practicable – for each state as initial state, a quench was done, i. e. only downward moves in energy were allowed until a local minimum was reached and then the resulting energies were compared. With two or more different results, the system was known to feature several local energy minima. Another method, when the state space is too large to be searched in full, would be to choose some random initial states and do a quench for them. If that leads to different resulting energies, the system has more than one minimum, else it *probably* doesn't. By the way, this quench technique can also be used to search for the global optimum and has been in age-long rivalry to the simulated annealing technique in what concerns effectiveness, that is: serving with fast results.
Choosing the J_{ij} from $[-1;\ 1]$ also accounts for $\bar{E}\,(T \to \infty) = 2^{-L} \sum_\alpha E_\alpha \approx 0$.

4.3 State Numbering For State Space Exploration

For use with the temperature algorithm, one needs all the states to determine $\underline{\underline{\Pi}}$ and all the state energies to calculate the objective function value $f = \underline{\mathrm{E}}\ \underline{\mathrm{p}}$.[1] Therefore, the states were numbered in the following way: Aligning all the spins in a chain, one starts with all spins set to -1 as state number 1. Then, flipping the last spin in the chain to 1, one obtains state 2. One continues to flip the last possible spin so as to create a new state and then flips the last 2 possible spins to 1 compared to state 1 and so on, until the state with number 2^L is reached where all the spins s_i are 1.

For the example of 3 spins, the state numbering is rendered in table 4.1.

[1] See section 6.2

State	Spins		
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	1	-1	-1
5	-1	1	1
6	1	-1	1
7	1	1	-1
8	1	1	1

Table 4.1: States For 3 Spins

5 Move Class Models

This chapter is concerned with modifications to the move class. To minimize misunderstandings, first of all a couple of naming conventions will be made, relying on the names already defined in section 2.2. For a short reference see also the glossary appended. The second section describes a most "natural" generalization of the neighborhood concept, whereas the third section presents two more approaches – trying to erase some rather unwanted effects of the first model. The fourth section presents sample neighborhoods for the example of 3 spins.

5.1 Nomenclature

In order to find a rule for dynamical adjustment of the underlying move class of the random walk, one first has to think about reasonable modifications to the *fundamental neighborhood relation* giving the set of all $\alpha \in N^1(\beta) \equiv N(\beta)$. How that is defined in detail is a matter of the specific system and the used move class model, but normally it is composed of the nearest neighboring states in state space.
Practically, the random walker makes a *fundamental choice* of a prospective succeeding state α from among the fundamental neighbors of a current state β. Then, the acceptance rule is applied to the difference in energy between the two states. Such a move will hence be called a *fundamental move.*

Wishing for faster convergence, one may try to increment the fundamental move class by defining 1 move as comprised of several fundamental choices and applying the acceptance decision afterwards. Such a move will hence be referred to as an *extended move*, and, more precisely for m fundamental choices, a move of order m.
The *move order* m is being introduced here for easier distinction between the different move lengths as a natural number giving the maximum number of fundamental choices allowed per extended move. Consequently, the fundamental move class is that of order $m = 1$.

For order m, equation (3.2) for the total transition probabilities becomes

$$\Gamma_{\alpha\beta}^{(m)}(T) = \Pi_{\alpha\beta}^{(m)} \times P(\Delta E, T), \tag{5.1}$$

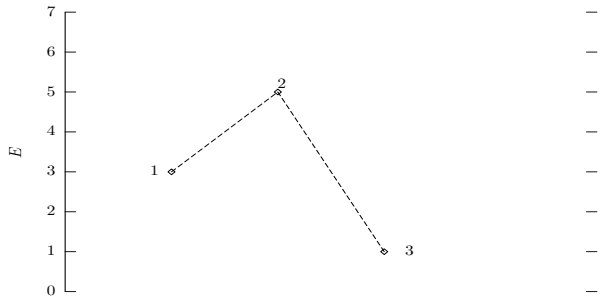

Figure 5.1: A simple three state barrier system

where α now is an $m^{\,th}$ *order neighbor* of β :

$$\alpha \in N^{(m)}(\beta).$$

Accordingly, the master equation is written as

$$\underline{p}^{\,t+1} = \underline{\underline{\Gamma}}^{(m)} \underline{p}^{\,t}.$$

Now, consider the example of $m = 2$, i. e. of making two fundamental choices before applying the Metropolis rule. The random walker would then choose an α from among the next neighbors of its own next neighbors ζ, $\zeta \in N^1(\beta)$:

$$\alpha \in N^2(\beta) = N^1\left(N^1(\beta)\right) = N^1(\zeta).$$

Practically, it would first choose a first-order neighbor ζ and from there again choose a nearest neighboring state $\alpha \in N(\zeta)$, a 2^{nd} order neighbor of β. Then, the Metropolis acceptance rule would be evaluated with $\Delta E = E_\alpha - E_\beta$. Leaving out the probability weighting for the intermediate positions significantly flattens the energy landscape! Figure 5.1 shows a simple three-state energy landscape. While it would be rather unlikely at finite temperature for the random walker to move *upwards* in energy along the bond from state 1 to 2 in order to be able to descend to the global minimum at 3 by making two separate fundamental moves, with such a 2^{nd} order move as described above, the walker is able to move *downward* directly from 1 to 3, the acceptance probability for which is 1.

What is needed to determine the structure of $\underline{\underline{\Pi}}^{(m)}$ is the respective neighborhood definition, obviously. Defining a *natural* fundamental *neighborhood relation*, the fundamental neighborhood is comprised of all the states α directly adjacent to state

β in state space, where $\beta \neq \alpha$: For continuous state spaces, these are the states within an interval of width $|\underline{\Delta p}|$ and $|\underline{\Delta q}|$ around the current state. For the discrete case, these are the states which differ from the current one by a fundamental change in configuration; with respect to the systems considered here[1], such a fundamental change would be the flip of 1 spin. A choice or move with underlying natural move class will be called a *natural choice or move* from now on.
Apart from the most simple-hearted approach, producing a chain of exactly m natural choices leading out from a starting state for an extended move of order m presented in the next section, other possible definitions allowing for up to m naturals per extended move are given thought to in the third section of the chapter at hand. To differentiate between the models, m is given an index for each in turn.

5.2 Making Exactly m Natural Choices Per Extended Move

This model adopts the natural neighborhood as its fundamental one, which for order $m = m_e$ is simply taken to its m^{th} power.

For given move order $m = m_e$, the random walk will be constituted of extended moves each consisting of m natural choices done with probability $1/|N_e|$. Or, expressed by the neighbor choice matrix:

$$\left(\underline{\underline{\Pi}}^1{}_e\right)_{\alpha\beta} = \frac{1}{|N_e^1(\beta)|}.$$

Following equation (3.3), this leads to

$$c = 1/|N_e^1(\beta)|$$

and all the diagonal entries of $\underline{\underline{\Pi}}$ are zero: A state is not a member of its own fundamental neighborhood.
Being in state β a next state α is chosen such that α is an m^{th} order natural neighbor of β:

$$\alpha \in N_e^{(m_e)}(\beta) = N_e^{m_e}(\beta).$$

This is done by making m fundamental choices after one another. The procedure would be, for example, to generate a natural number $\xi \in [1,\ |N_e(\beta)|\ ]$ and then choose the $\xi^{\text{ th}}$ among the fundamental neighbors. For spin systems that means, that the $\xi^{\text{ th}}$ spin s_ξ would be flipped to $-s_\xi$. This has to be done m times and

[1] See chapter 4.

only after choosing, the Metropolis rule is to be applied to the finally chosen α and starting state β.

For order m_e, equation(5.1) for the total transition probabilities looks like this

$$\begin{aligned}\Gamma_{\alpha\beta}^{(m_e)}(T) &= \left(\underline{\underline{\Pi}}_{e}^{m_e}\right)_{\alpha\beta} \times P(\Delta E, T) \\ &= \underbrace{\left(\underline{\underline{\Pi}}_{e}\,\underline{\underline{\Pi}}_{e}\cdots\underline{\underline{\Pi}}_{e}\right)_{\alpha\beta}}_{m_e \text{ factors}} \times P(\Delta E, T).\end{aligned}$$

It must be noted here, that of course $\left(\underline{\underline{\Pi}}_{e}^{m}\right)_{\alpha\beta} \neq 1/|N_e^1(\beta)|$. m^{th} order neighbors are no longer chosen with equal probability. What's more, with increasing m_e, the number of non-zero entries in $\underline{\underline{\Pi}}$ increases. Here also a remarkable effect appears: Since the fundamental choices are done after each other, the random walker can also move backwards in the chain of intermediate states. Such a backward turn involves two steps, one there and one back to the original position. An m^{th} order choice with $m_e = 2k,\ k = 1, 2, 3, \ldots$ on the one hand can be depicted as itself being comprised of k 2^{nd} order choices and can thus lead either 0 or 2 or $4 \ldots$ or m_e steps away in state space in summary. An m^{th} order choice with $m_e = 2k + 1$ on the other hand is comprised of such an even-m-choice plus one natural choice so that it can lead either 1 or $3 \ldots$ or m_e steps away from the starting position in state space. Like the fundamental move class, a random walk of such extended uneven order moves may reach anywhere within the state space in a due number of steps. Contrarily, for even m, this neighborhood definition is *not generally ergodic*! That means, not every state can be reached from every other. Because one 2^{nd} order choice can only either end at the original position or two steps away from it, the random walker will never be able to reach a position which is an uneven number of choices away from the origin– at least if the origin itself cannot be reached by an uneven number of choices. This in turn highly depends on the state space structure. However, for spin systems, it obviously divides the state space into two halves for a given starting point β; one containing the even- and one containing the uneven-order neighbors of β. The effect also results in a maximum of 50% non-zero entries in $\underline{\underline{\Pi}}^m$.

But consider a closed chain of 5 states, the random walker may move around on: Here, each state is an even- and uneven-order neighbor of itself at the same time and consequently, all move classes of order m_e are ergodic.

For very large configuration spaces in practice the differentiation might be rather impossible, so that generally only uneven m_e should be used. Note that for a *move* of any order at finite temperature, there is always a probability to stay in the current

state, yet that does *not* contribute to ergodicity. Anyway, other models based on that at hand were designed to avoid this issue right from the beginning.

5.3 Allowing Up To m Natural Choices Per Extended Move

In the following, two approaches avoiding non-ergodicity will be presented.

5.3.1 A Model With Inherent Inertness

This model uses a modified fundamental neighborhood definition as basis. Again, order $m = m_i$ means just to take the m^{th} power of the fundamental quantities.

Considering equation (3.3) for the diagonal elements, $\Pi_{\beta\beta}$ obviously depends on c. When choosing $c \neq 1/|N_e^1(\beta)|$, the diagonal entries do not vanish any more and there is always a probability for the system not even to *attempt* to move anywhere else; even for infinite temperature

$$\beta \in {N_i}^1(\beta)\,.$$

In a somewhat physical interpretation, c thus turns out a constant specifying something like the system's inertness or time behavior, its inherent motivation for leaving its current state. A state β can now be chosen as its own successor, i. e. it can be reached by an uneven number of fundamental choices, in this way nullifying the objections made on ergodicity in the last section.[2]

In detail, $c = c_i$ was here chosen as

$$c_i = \frac{1}{|N_e(\beta) + 1|} \equiv \frac{1}{|N_i(\beta)|}$$

so that all non-zero entries $\left(\underline{\underline{\Pi}}^1_i\right)_{\alpha\beta} = 1/|N_i(\beta)|$ equally. Again, for move order $m = m_i$, the inert $\underline{\underline{\Pi}}^1_i$ is taken to the m^{th} power:

$$\underline{\underline{\Pi}}^{(m_i)}_i = \left(\underline{\underline{\Pi}}_i\right)^{m_i},$$

and then the Metropolis rule is applied to obtain the transition probabilities according to (5.1). As for the random walk association, one would generate a natural random number $\xi \in [1,\ |N_i(\beta)|]$ uniformly distributed and if $\xi = 1$, stay in the current state,

[2] This effect appears also for degenerated systems, where there are many β and transitions between them.

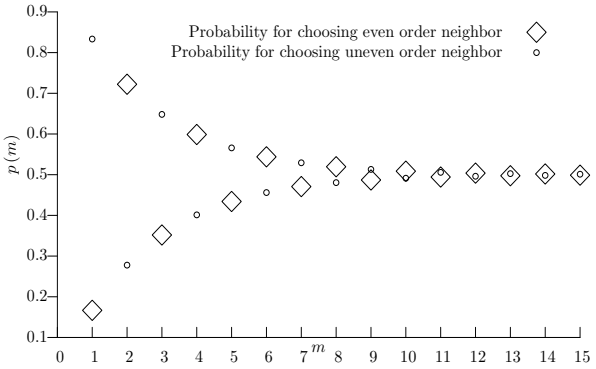

Figure 5.2: Evolution of probabilities for choosing even and uneven order neighbors

else choose another nearest neighbor analogous to the procedure depicted in the last section, only that the $(\xi + 1)$ $^{\text{th}}$ neighbor is now chosen for a generated ξ.
For an extended move, this is repeated m times and the cost function values of the starting and resulting positions are then used for decision whether to accept the choice made. That means, that generally for a move of order m, the random walker may end in a position 0 or 1 or 2 or... or m steps away from the starting point – only the probabilities differ: For small even m_i, to choose an even (up to m^{th}) order neighbor is much likelier than to choose an uneven order one. For the example of 5 spins, the evolution of probabilities is plotted in figure 5.2. Each state has 5 natural neighbors. For $m_i = 2$, the first choice leads with probability 5/6 to one of these; with 1/6, the random walker stays where it is. For the second choice, again the random walker chooses to go to a natural neighbor in 5/6 of all cases and only stays in 1/6. In summary, the random walker will choose an even order neighbor, i. e. either its starting point or a neighbor two steps away with probability $p = 1/6 \times 1/6 + 5/6 \times 5/6 = 13/18$, and only with 5/18 chooses a uneven order, natural neighbor of its starting point. Therefor, although the move class is in principle ergodic, especially for small even m, effects relating to the ood-even-phenomenon concerning ergodicity will be observed, as shown in the results' part.

5.3.2 A Model For Choosing Different Orders

However, another model was investigated, neglecting inertness and leaving the diagonal entries for the fundamental move class to be 0. Instead, the fundamental move class is again the natural one, yet for order m, it is no longer taken to the m^{th} power,

but the neighbor choice probability matrix $\underline{\underline{\Pi}}_c^{(m)}$ was taken to be the sum of all $\underline{\underline{\Pi}}_e^k$ of lesser order divided by $m = m_c$, making the choice of extended moves of order up to m_c equally likely with probability $\frac{1}{m_c}$.

$$\underline{\underline{\Pi}}_c^{(m)} = \frac{1}{m_c} \sum_{k=1}^{m} c\underline{\underline{\Pi}}_e^k$$

In that way,

$$\underline{\underline{\Pi}}_c^{(1)} = \underline{\underline{\Pi}}_e{}^k.$$

Since the column sums of each constituting summand are equal to 1, the total column sum for m summands is m and division by m brings $\underline{\underline{\Pi}}_c^{(m_c)}$ in accord again with equation (3.3). Practically, the random walker would first generate a natural number $\xi \in [1,\ m_c]$ before each oncoming move to choose the move order to use. Then it would jump to a next state by first making exactly $\xi = m_e$ natural neighbor choices as depicted in section 5.2 and afterwards applying the acceptance rule. For chosen order ξ even or uneven, the random walker is again restricted to neighbors an even or uneven number of steps away, respectively. But since ξ is chosen anew before every move, ergodicity can be assumed nonetheless.

5.4 Layout Of Spin Neighborhoods

For spin systems, each state β has L natural next neighbors α – arrangements $\{s_i^\alpha\}_{i=1,\ldots,\ L}$ where precisely 1 spin is flipped to $s_j^\alpha = -s_j^\beta,\ \alpha = j = 1, 2, \ldots,\ L$. The number of neighbors with exactly m spins flipped compared to a certain alignment is easily to be determined with combinatorial considerations as binomial coefficient of L over m. Therewith, for order $m_e \leq L$, the number of neighbors is:

$$|N_e^m| = \begin{cases} \displaystyle\sum_{j=0}^{(m-1)/2} \begin{pmatrix} L \\ 2j+1 \end{pmatrix} & \text{for } m \text{ uneven} \\ \displaystyle\sum_{j=0}^{m/2} \begin{pmatrix} L \\ 2j \end{pmatrix} & \text{for } m \text{ even} \end{cases}$$

For $m_e > L$, $|N_e^m|$ remains constant at 2^{L-1} for both even and uneven m. Here was utilized that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ [3] and $\binom{n}{k} = \binom{n}{n-k}$ wherefor

$$\sum_{j=0}^{(L-1)/2} \binom{L}{2j+1} = \sum_{j=0}^{L/2} \binom{L}{2j} = 2^{L-1}.$$

Since the fundamental move class of the inert model includes the possibility of no spin flip, all the possibilities for flipping exactly m spins have to be added up for order m_i. The number of neighbors for $m_i \leq L$ is:

$$|N_i^m| = \sum_{k=0}^{m} \binom{L}{k}.$$

and for higher $m > L$ has the constant value of 2^L. Then, each state is each other's neighbor and $\underline{\underline{\Pi}}$ has no more non-zero entries.

For order m_c, the neighborhood contains

$$|N_c^m| = |N_e^m| + \left|N_e^{m-1}\right|$$

m_{th} order neighbors: Due to the possibility that the random walker steps backwards or makes a loop in the choice chain, $N_e^{m-2} \subseteq N_e^m$ has to be heeded. No restriction has to be made to the value of m_c because the maximum of $|N_e^m|_{\max} = 2^{L-1}$.

With the state sequence as given in table 4.1, all the natural neighbors of a state β are given by the following piece of Mathematica code:

```
For[k=1, k<=L, k++,
    connections[[k]]=
      Table[If[Mod[i,2^k,1]<=2^(k-1), {i,i+2^(k-1)}->1,
            {i,i-2^(k-1)}->1], {i,1,2^len}]];
```

That means, depending on the value of `Mod[i,2^k,1]`, either states `i and i+2^(k-1)` or states `i and i-2^(k-1)` are connected as natural neighbors. Again, for the ex-

[3] $(1+1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} 1^k = 2^n$

ample of 3 states, the natural $\underline{\underline{\Pi}}_e$ then looks like:

$$\underline{\underline{\Pi}}_e^1 = \frac{1}{3}\begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

For $m_e = 2$, one can easily see the non-ergodicity from the structure of $\underline{\underline{\Pi}}$ which is here shown again for the example of 3 spins:

$$\underline{\underline{\Pi}}_e^2 = \frac{1}{9}\begin{pmatrix} 3 & 0 & 0 & 2 & 0 & 2 & 2 & 0 \\ 0 & 3 & 2 & 0 & 2 & 0 & 0 & 2 \\ 0 & 2 & 3 & 0 & 2 & 0 & 0 & 2 \\ 2 & 0 & 0 & 3 & 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 & 3 & 0 & 0 & 2 \\ 2 & 0 & 0 & 2 & 0 & 3 & 2 & 0 \\ 2 & 0 & 0 & 2 & 0 & 2 & 3 & 0 \\ 0 & 2 & 2 & 0 & 2 & 0 & 0 & 3 \end{pmatrix}$$

By comparing with table 4.1, one may easily verify that 0 or 2 spin flips are allowed. The non-zero-entries have also already reached their maximum at 50% of all elements of $\underline{\underline{\Pi}}$. Also, starting in a certain state, e. g. state 1, one cannot reach everywhere: There is a transition possibility to states 1, 4, 6 and 7. From state 4, one can only reach 1, 4, 6 or 7 and so on, so that the four states 1, 4, 6 and 7 are separated from the four other states! Now, compare this observation to that for $m_e = 3$:

$$\underline{\underline{\Pi}}_e^3 = \frac{1}{27}\begin{pmatrix} 0 & 7 & 7 & 0 & 7 & 0 & 0 & 6 \\ 7 & 0 & 0 & 7 & 0 & 7 & 6 & 0 \\ 7 & 0 & 0 & 7 & 0 & 6 & 7 & 0 \\ 0 & 7 & 7 & 0 & 6 & 0 & 0 & 7 \\ 7 & 0 & 0 & 6 & 0 & 7 & 7 & 0 \\ 0 & 7 & 6 & 0 & 7 & 0 & 0 & 7 \\ 0 & 6 & 7 & 0 & 7 & 0 & 0 & 7 \\ 6 & 0 & 0 & 7 & 0 & 7 & 7 & 0 \end{pmatrix}.$$

Both 1 or 3 flips are allowed here, $\underline{\underline{\Pi}}$ has no diagonal entries and what concerns ergodicity: From state 1, the random walker can go to 2, 3, 5 or 8. Yet from state

2, 4, 6 and 7 can also be reached. In that way all the state neighborhoods are interconnected.
For the inert move class model, the fundamental neighborhood choice matrix is:

$$\underline{\underline{\Pi}}^{1}_{i} = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

For $m_i = 2$ already, each state has 7 neighbors, only the state where all 3 spins are flipped in the opposite direction being excluded, and from $m \geq L = 3$, there are no more zero-entries, i. e. each state is each other's neighbor.
For the model, where the order m_e to use is chosen before each move, the fundamental $\underline{\underline{\Pi}}^{1}_{c} = \underline{\underline{\Pi}}^{1}_{e}$. And for $m_c \geq 2$ the structure is the same as that for $\underline{\underline{\Pi}}^{2}_{i}$. Again, for $m_c = m \geq 3$, each state is each other's neighbor.

6 The Plan

6.1 The Idea With The Correlation

As hinted at above, the effect of increasing the move order is to flatten the energy landscape, thus making it easier for the random walker not to get stuck in a local minimum but to find the global minimum. Unfortunately, the probability for staying in or near the ground state also is lessened when allowing the walker to happily hop around in state space wearing "seven-league boots." What is needed is an indicator enabling the walker to decide which boots to put on, i. e. how long steps to make. This can differ during an annealing run, probably the optimal move order changes with temperature.
Considering standard statistical functions, what could be more manifest than have a look at the correlation of energies at different stages of an annealing run? Its pattern might show, after how many steps the random walker has no more or sufficient little knowledge about its starting point. Yet, it depends on the temperature. But, since one needs to decide about the memory dependance on the number of *choices* without the Metropolis probabilities rather than moves, it seems the best to consider the case of high T, especially $T \to \infty$ more closely.
The correlation function is defined as

$$r(k) = \frac{< E_0 - \overline{E} >< E_k - \overline{E} >}{\sqrt{< \left(E_0 - \overline{E}\right)^2 >< \left(E_k - \overline{E}\right)^2 >}}, \tag{6.1}$$

where $\overline{E}$ is the overall arithmetic energy mean – already boltzmannized by application of the Metropolis rule. Starting in an initial state with E_0, the random walker proceeds to others with energies $E_k, k = 1, \ldots, k_{\max}$ by making k fundamental moves, fundamental because the move class is not yet adjusted. The procedure would be to generate a random initial state with energy E_0, let the walker do $k_{\max}$ moves while storing the energy chain $E_1,\ E_2, \ldots,\ E_{k_{\max}}$ in an array, then leave out several moves to obtain a bit of statistical independence and take the next energy as a next initial from which again a chain of length $k_{\max}$ is recorded, and so on for a couple of series. Now, $r(k)$ can be computed from the data collected – assume that there was no significant change in temperature in the meantime, because $\overline{E}$ is the mean

energy of the system and thus subject to temperature change due to the Metropolis rule. Obviously, $r\,(0) = 1$ and decreases for greater k, its slope depending on the temperature. For infinite T, $r\,(k) \to 0$ for rising k because all moves are accepted – quickly brainwashing the walker's memory of preceding positions. For finite T, $r\,(k)$ will decrease rather gently and for small T, approach a level above zero because then moves upward in energy become very unlikely – the random walker is more or less only pushed downwards from its original height, and if once caught in a minimum won't easily be able to leave it again. Finally, for $T \to 0$, $r\,(k)$ will settle at 1, because the random walker will never be able to leave a minimum again. The run of the curve, however, hopefully exhibits a characteristic length b when fitted to an exponential function of the form $e^{-b\,k}$ which shows an interrelation with the optimal move order. For computation of $r\,(k)$, the fundamental natural move class will be used only, as for $m = 1$, this is the fundamental both for $m_{\,e}$ and $m_{\,c}$ anyway, and, as will be seen in the results' chapter, $m_{\mathrm{opt,\,i}} = m_{\mathrm{opt,\,e}}$.

6.2 Affordable System Sizes And Algorithm Parameters

The temperature schedule algorithm of section 3.2 was implemented in Mathematica®, which was chosen for its ability to handle sparse matrices necessary due to the implementation of the temperature schedule algorithm: To speed up computation the transition matrices $\underline{\underline{\Gamma}}^{(m)}$ were precomputed for specified values of temperature steps: The temperature $T \in [\,0;\,+\infty]$ is scaled to $x = e^{-1/T} \in [\,0;\,1]$, divided into 100 intervals, that means 101 transition matrices had to be precomputed and stored. With, e. g. 7 spins, that produces 101 128×128-matrices $\underline{\underline{\Gamma}}^{(m)}\,(x\,)$, fortunately for limited memory sizes, most of them having a lot of zero-entries due to restricted neighborhood relations. As initial distribution the stationary distribution for infinite temperature was chosen, for these simple systems: ${p_\alpha}^0 = 1/(\#\ states)$. The main algorithm constants applied are shown in table 6.1. Also, other constants were used

Name	Meaning	Value
div	number of temperature intervals	100
nmax	number of annealing steps	100
accur	accuracy for determination of temperature schedule algorithm	10^{-5}

Table 6.1: Algorithm constants for annealing

for comparison but neither $n = 1000$ nor $div = 1000$ or $accur = 10^{-8}$ did have any influence on the quality of the results with respect to m.

The schedule determination algorithm did already exist and was here used as a black box only, with input: Transition matrices and state energies. The energies were needed for the objective function, as explained below and the transition matrices had to be computed for each respective move class model and order.
The objective chosen is to minimize the mean final energy of an annealing run:

$$f = \underline{E}^{\mathrm{T}}\, \underline{p}^{\,n},$$

where $\underline{E}$ is a vector containing all the state energies in concordant order with that in the final probability distribution vector $\underline{p}^{\,n}$. The final objective function value then is $f^{\,n} = \overline{E_f}$ and is to be understood as the mean final energy one would obtain when doing a couple of real annealing runs. Another often employed objective is to maximize the final probability of being in the ground state in which case $\underline{E}$ has all zero entries except for the ground state position: $E_{\mathrm{GS}} = -1$.

In summary, on a macroscopic scale, only very small state spaces can be examined because

- one needs exhaustive knowledge about the system's state space: All the neighborhood relations are needed for $\underline{\underline{\Pi}}$ and all the energies for $\underline{E}$.
- the transition matrices need yet a considerable amount of memory and computation time: For the example of 7 spins, there have to be $128 \times 128 \times 101$ entries computed.
- the computation time increases with move class order: Since backward steps in the choice chain are allowed, the number of possible neighbors and thus the number of non-zero entries in $\underline{\underline{\Pi}}^{(m)}$ increases significantly with m which on its part leads to more non-zero entries in $\underline{\underline{\Gamma}}^{(m)}$.

The absolute maximum size for spin systems to be handled in acceptable time is $L = 9$, for which the computation for only one m-value lasts already about 24 hours on a conventional x86 PC – for the university's compute servers, this time is about half the amount, depending on their current workload of course. Therefore, for $L \geq 7$, only $m = m_e$ and uneven will be investigated in general. As $\underline{\underline{\Pi}}_{\,e}$ has a maximum of 50% non-zero entries, the computation time for this model is generally lower than for the other two, where the number of non-zero entries decreases fast with rising m. For an example, see section 5.4. The characteristics modified for systems of different sizes are rendered in table 6.2.

The correlation $r\,(k)$ was computed in detail in the following way: Starting with a randomly generated state with $E\,(0)$ – the choice of any of which is equal to that of the

Name	Description	Range
L	Characteristic Size	5-8
m	Move class order	1-7
h	Hamming distance between local minima	$2 - L$

Table 6.2: Modified characteristics for annealing

others – a series of energies $E(k)$ is created by an annealing run, for $k \leq k_{\max} = 15$. Then, a number of 20 moves is left out of the calculation, and the 21^{st} energy after the end of the preceding series is taken as $E(0)$ of the next series. In that way, an array of $E(k)$ is produced for an amount of 1000 series. Then, $r(k)$ is computed following equation (6.1). After that, an exponential fit is made for $r(k)$, adjusting the parameter b by a linear least squares method.

7 Detailed Results

This chapter will present selected results obtained for spin systems of different size and number of energy minima. The first section of this chapter presents some common features of the different move class models and the correlation pattern $r(k)$ in detail for the example of a system with only 5 spins. Then, other spins systems of greater size are investigated with respect to the obtained optimal move order and the computed characteristic length b of the correlation. The Appendix contains all the state space characteristics and results for all sample systems in detail.

7.1 Common Features Shown For The Example Of 5 Spins

The system considered in detail is system 5a. System 5a has 3 energy minima, the global minimum of which has a potential energy of -4.65509. The other two minima can each be reached from the global one by 3 spin flips; in other words, they are three natural fundamental steps in state space away.

7.1.1 Temperature Schedule Pattern

Figure 7.1 shows a typical pattern of an optimal temperature schedule determined for the natural move class $m_e = 1$. The concrete temperature schedules depend on the underlying neighborhood model and move class order, of course. But the temperature range for the shown schedule is typical for all systems: As the number of spins does not differ much for the different system sizes, the maximum energy heights to be climbed by the random walker also don't. But that is just what the temperature range depends on. With rising m, the whole schedule becomes less continuous, with sudden drops down $T = 0$ or jumps up to $T \to \infty$, until the whole schedule finally turns into "quench" for $m \gtrapprox L$. Also, for even m_e, the schedule sometimes already says quench but for higher, uneven m_e it doesn't – a result of the non-ergodic underlying move class.

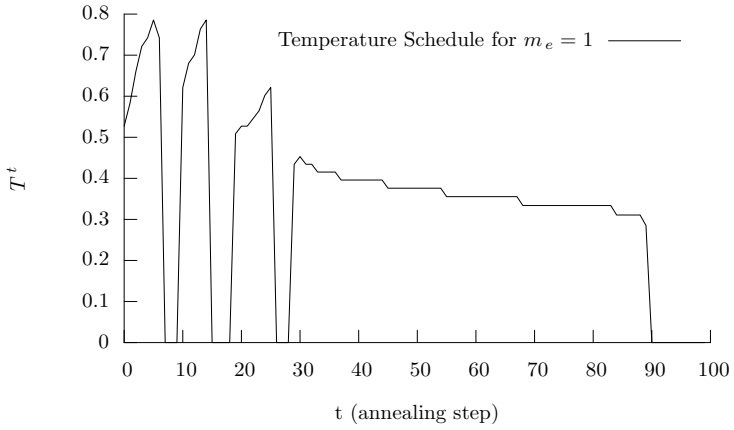

Figure 7.1: Temperature schedule for $m_e = 1$ for system 5a.
The pattern and temperature range is typical for all the systems investigated.

7.1.2 Features Of Different Move Class Models

Figure 7.2 summarizes the results obtained for system 5a and different move class models for different move orders. It shows the oscillatory behavior of the computed mean final energy $\overline{E_f}(m)$ [1] due to non-ergodicity for even m_e as well as the slighter oscillation for small even m_i due to the higher probability of the choice of an even order natural neighbor. Remarkably, for $m_e = 2$, $\overline{E_f} = -4.242$. As pointed out, for even move order, model m_e is not ergodic for spin systems. Since the chosen starting probability distribution of states is chosen as uniform, the states of each half of the state space have a total probability of 0.5. When taking the absolute minimum energies of both separate halves of the state space for this system and weighting these two values with probability 0.5 each, one obtains (almost) exactly that very result. This is the clearest indication for the correctness of the theory one could wish for. Also, it becomes clear that small uneven m_e yield better results than m_i, because of the inherent ability there not even to attempt a move; and these in turn are better than the results for m_c, only for $m = 1$, the results for m_e and m_c are the same – as should be, looking at the definition of the move class models.

[1] Note that the final objective function value $f^n = \overline{E_f}(m)$.

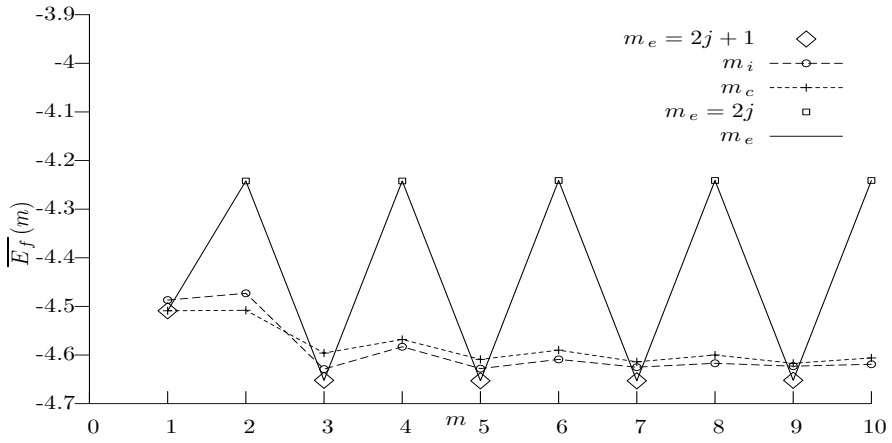

Figure 7.2: Resulting mean final energy for system 5a and different move class models. The oscillatory behavior for even $m = m_e$ is due to non-ergodicity. Yet, uneven $m_e \leq L$ yield the best results.

7.1.3 The Correlation Pattern

Figure 7.3 shows the standard deviation for 10 computations of b for system 5a for different temperatures. For this, $r(k)$ was computed 10 times in the usual way, and then the arithmetic means were taken as $\overline{b}(T)$. Their deviations are shown as error bars. Because for low temperatures the random walker is likely to get stuck in a local minimum at some point, the computed $r(k)$ show a high relative standard deviation: For $T = 0.2, \Delta b/ \approx 0.9$. For higher temperature, the relative standard deviation of the determined characteristic lengths settles at a level of about $0.05 - 0.1$. Thus, for comparison with m_{opt} mainly the more reliable $b(T \to \infty)$ should be used.

Figure 7.4 shows the pattern of $r(k)$ fitted to an exponential function for different temperatures. Apart from the fact, that the exponential fit appears to be a good choice, it can be clearly seen, that with rising temperature, the slope of $r(k)$ increases, as expected.

Figure 7.5 compares the temperature dependency of the fit parameter b for underlying fundamental $\underline{\underline{\Pi}}_e$ and $\underline{\underline{\Pi}}_i$. The overall pattern is a decrease to about the same value for $T \to \infty$, yet for moderate temperatures, there are clearly differences: For model $m_i = 1$, the correlation between energies decreases faster from $b(0) = 1$ for both models, and then shows a gentler slope than that for $m_e = 1$ – this is due to the higher affinity of the random walker to stay where it is.

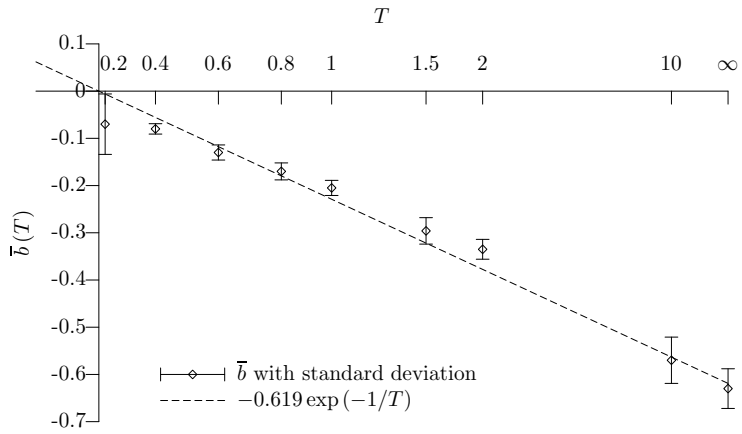

Figure 7.3: $b\,(T)$ with standard deviation for system 5a.
For low temperatures $T \lessapprox 0.4$, the computed characteristic length b is not very reliable.

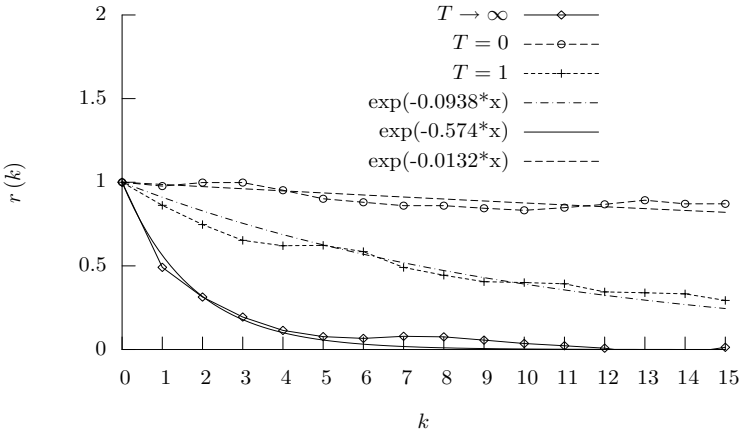

Figure 7.4: $r\,(k)$ for different temperatures for system 5a.
The slope of the energy correlation increases with rising temperature, as expected. The exponential fit to the correlation pattern is the right choice.

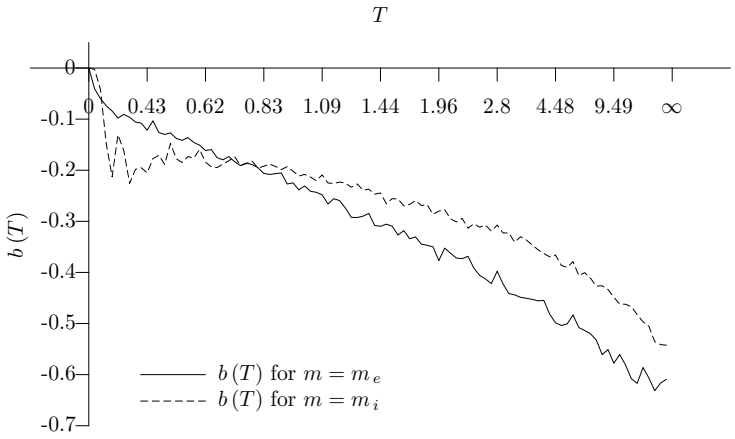

Figure 7.5: $b\,(T)$ for different fundamental neighborhoods.
Because of the always present probability for the inert system not even to attempt a move, it shows a gentler slope for $b\,(T)$.

7.1.4 Conclusions For Further Investigations

Since in general uneven $m_e \leq L$ account for the best results, only these will be used for the larger systems of 7, 8 and 9 spins – due to limited time for computation. Also, $r\,(k)$ will be computed for the natural fundamental move class only, because for the same systems, m_{opt} for models m_e and m_i are the same. As characteristic lengths $b\,(T)$ for $T = 0.4$ and $T \to \infty$ will be used. m_{opt} is determined with respect to the computational effort of a move, which rises with m. Thus, if the results are very close, the lower m is taken as optimal.

7.2 Systems With 5 Spins

This and the following sections summarize the results obtained when trying to find a rule to determine m_{opt} by computing the energy correlation during an annealing run and fitting it to an exponential function of the form $e^{-b\,x}$ for sampling points $(k; r\,[k])$. To obtain somewhat reliable results several systems with 5, 6, 7, 8 and 9 spins were investigated – possessing different numbers of energy minima with different Hamming distances and intensities.

The characteristics for the systems with 5 spins are given in table A.1 in the appendix. In appendix B, table B.1 yields all the computed mean final energies for

$m \leq L = 5$. There, the respective optimal results are marked by a *. The optimal move order shows a clear dependency on the Hamming distance between the local and global minima. Yet, for systems b, c and d, with an even Hamming distance of the two minima, an adjacent uneven m_e is best – just another validation for non-ergodicity of even m_e. Also, the systems considered only have 1 *local* energy minimum and thus only 1 characteristic h. For systems with more minima and different h, these dependencies interact and there is no clear interrelation between h and m_{opt} any more.

Tables 7.1 and 7.2 below summarize the optimal move orders m_{opt} and the exponential fit parameters b for temperatures $T = 0.4$ and $T \to \infty$ for m_e and m_i, respectively. Since the systems are similar in size, measured at the same temperature, effects on $r(k)$ relating to these two parameters have no influence. Though there is no *clear* dependency to be observed, for system 5e, where $m_{\text{opt}} = 5$, b is highest.

System	5a	5b	5c	5d	5e
$m_{\text{opt},\,e}$	3	3	3	3	5
$b_e\,(T = 0.4)$	0.094	0.093	0.19	0.12	0.20
$b_e\,(T \to \infty)$	0.57	0.68	0.64	0.77	0.85

Table 7.1: $m_{\text{opt},\,e}$ and $b\,(T)$for 5-spin systems.
There is no dependency to be observed between the optimal move order and the characteristic length of the energy correlation.

System	5a	5b	5c	5d	5e
$m_{\text{opt},\,i}$	3	3	3	3	5
$b_i\,(T = 0.4)$	0.016	0.17	0.033	0.0077	0.018
$b_i\,(T \to \infty)$	0.58	0.56	0.61	1.15	0.65

Table 7.2: $m_{\text{opt},\,i}$ and $b\,(T)$for 5-spin systems.
There is no dependency to be observed between the optimal move order and the characteristic length of the energy correlation.

7.3 Systems With 6 Spins

Several systems with 6 spins featuring different characteristics given in table A.2 were investigated. The results are given in table B.2. Again, the results obviously depend

on the Hamming distance between the minima, but as tables 7.3 and 7.4 show, there is no dependency to be observed between the m_{opt} and the fit parameters b unless it is, that b rises slightly with the optimal move order.

System	6a	6b	6c	6d
$m_{\text{opt},\,e}$	5	5	5	3
$b_{\,e}\,(T = 0.4)$	0.10	0.088	0.069	0.085
$b_{\,e}\,(T \to \infty)$	0.73	0.70	0.54	0.66

Table 7.3: $m_{\text{opt},\,e}$ and $b\,(T)$for 6-spin systems.
There is no dependency to be observed between the optimal move order and the characteristic length of the energy correlation.

System	6a	6b	6c	6d
$m_{\text{opt},\,i}$	5	5	5	2
$b_{\,i}\,(T = 0.4)$	0.056	0.13	0.0061	0.15
$b_{\,i}\,(T \to \infty)$	0.58	0.74	0.40	0.48

Table 7.4: $m_{\text{opt},\,i}$ and $b\,(T)$for 6-spin systems.
There is no dependency to be observed between the optimal move order and the characteristic length of the energy correlation.

7.4 Systems With 7 Spins

Table A.3 shows the characteristics of the investigated systems. Table B.3 gives the results for the different move classes. Figure 7.6 shows the results obtained for different neighborhood definitions and system 7a. $f^{\,n} = \overline{E_f}\,(m)$ is the mean final energy of annealing runs. Here, as for all the larger system of 7, 8 or 9 spins, the results become worse for $m \geq L$ until they approach a level: For $m \approx 20$, the series of $\underline{\underline{\Pi}}^{(m)}$ approaches a stationary neighbor choice probability matrix, with all entries equal. Therefor, the results for any move of higher order then are the same. Again, the results with only 2 or 3 local minima depend on the Hamming distances. Table 7.5 gives a final overview over the optimal move class order for $m = m_{\,e}$ and the exponential fit parameter $b\,(T)$ for low and high temperature. There is no clear interrelation and if any at all, b seems to be lower here for lower m_{opt} – in contradiction to the results for systems with 5 and 6 spins!

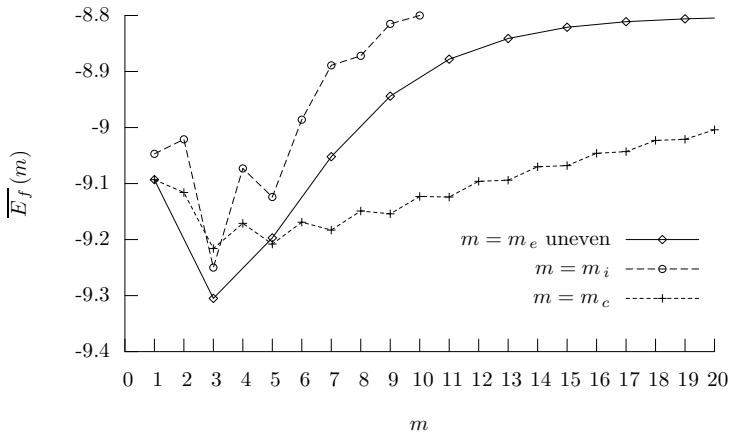

Figure 7.6: Mean final energy for 7 spins and different move class models.
For move order $m > L = 7$, the results become worse and then settle at a level as the neighbor choice probabilities reach a stationary distribution.

System	7a	7b	7c	7d	7e	7f
m_{opt}	3	3	1	1	3	1
$b\,(T = 0.4)$	0.040	0.14	0.16	0.096	0.055	0.11
$b\,(T \to \infty)$	0.47	0.50	0.52	0.48	0.44	0.52

System	7g	7h	7i	7k	7l
m_{opt}	1	5	5	1	3
$b\,(T = 0.4)$	0.11	0.091	0.014	0.044	0.16
$b\,(T \to \infty)$	0.57	0.52	0.53	0.50	0.44

Table 7.5: m_{opt} and $b\,(T)$for 7-spin systems.
There is no clear dependency to be observed between b and the optimal move order.

7.5 Systems With 8 Spins

Table A.4 shows the characteristics of the investigated systems. Like for 5, 6 and 7 spins, several systems with 8 spins featuring different characteristics given in table A.4 were experimented on. The results are given in table B.4. Table 7.6 gives a final overview over the optimal move order for $m = m_e$ and the exponential fit parameter $b(T)$ for low and high temperature. There appears to be no interrelation at all.

System	8a	8b	8c	8d	8e	8f
m_{opt}	1	3	3	1	1	1
$b(T = 0.4)$	0.097	0.068	0.084	0.051	0.055	0.22
$b(T \to \infty)$	0.46	0.48	0.42	0.41	0.43	0.44

System	8g	8h	8i	8k	8l	8m
m_{opt}	3	1	3	1	3	1
$b(T = 0.4)$	0.082	0.036	0.10	0.032	0.014	0.054
$b(T \to \infty)$	0.42	0.44	0.42	0.38	0.41	0.46

Table 7.6: m_{opt} and $b(T)$for 8-spin systems.The values show no interrelation.

7.6 Systems With 9 Spins

10 systems with 9 spins featuring different characteristics given in table A.5 were investigated. The results are given in table B.5. Table 7.7 gives a final overview over the optimal move class order for $m = m_e$ and the exponential fit parameter $b(T)$ for low and high temperature.

7.7 Conclusions

In general, uneven m are better than even m and only small $m <= L$, more precisely $m = 1,\ 3$ or 5 are optimal. These limits may be due to the small system sizes considered. But probably they relate to the rising number of neighbors possibly to be chosen with rising m which for finite temperature results rather in a random search than in annealing. The latter is also indicated by the fact that for system 5a, the results stay at their optimal level, whereas for 7 spins – with four times as many states – the results become worse with rising $m > m_{\text{opt}}$. Especially for m_e and

System	9a	9b	9c	9d	9e
m_{opt}	1	1	3	1	1
$b\,(T = 0.4)$	0.11	0.040	0.024	0.075	0.053
$b\,(T \to \infty)$	0.48	0.42	0.36	0.39	0.37

System	9f	9g	9h	9i	9k
m_{opt}	1	1	1	1	1
$b\,(T = 0.4)$	0.034	0.10	0.0088	0.10	0.013
$b\,(T \to \infty)$	0.38	0.40	0.40	0.40	0.37

Table 7.7: m_{opt} and $b\,(T)$for 9-spin systems.
There is no interrelation between the optimal move order and the characteristic energy correlation length.

m_i greater than L, the random walker seems to get rather astray. In a performance ranking, the first model with the natural neighborhood as fundamental one, $m = m_e$, is best. Then comes $m = m_i$ and after that only $m = m_c$, the latter of which has the advantage that small even m_c are more feasible than small even $m = m_e$ or $m = m_i$. Also, for higher m, the result stays at a lower level.

The energy surface must be sufficiently rough and the intensity of the local minima high enough for $m > 1$ to be good, because moves of order $m > 1$ significantly flatten the landscape. Thus, being in the global minimum region with energetically close metastable configurations, the system may easily leave that region for $T > 0$ to take on such a metastable state. This, in turn, contributes to the fact that the optimal schedules for high $m \gtrsim 5$ are simply a quench. Otherwise, the probability distribution of states would stay rather uniform instead of being changed to peaks at the local minima.

Another thing is, that for the special systems investigated, with rising system size one rarely finds any optimal move order than the fundamental $m = 1$. Other systems should be used to check whether this is due to the size of the system or merely a feature of these special Ising systems, as is believed by the author.

8 Summary

The results show no interrelation between the characteristic length of the energy correlation during an annealing and the optimal move order. It is likely, that for a precise decision about the move class to use one has to use another criterion, overlooked so far.

All the systems experimented on show the best results for the move class model with $m = m_e$, allowing for exactly m natural choices before applying the Metropolis acceptance rule, with the restriction that only uneven m should be used.

It may, however, well be that the optimal move class also depends on the temperature in such a way that at beginning of the annealing run, higher m are better to reach the region of the global optimum with greater speed, and then m has to be lowered to prevent the random walker from hopping out of that region again. Such a procedure would also nullify the non-ergodicity for a "pure even m_e run" and the effects relating to it for small even m_i.

Considering a real annealing run, a move of order $m > 1$ also needs more computation time than a fundamental move, dependant on the model, and it may well happen, that it needs even as much time as m separate fundamental moves. Therefore, the number M of moves allowed per annealing step and in total should be adjusted to the move class order, e. g. following $M(m) = \left\lceil \frac{M(1)}{m} \right\rceil$.

List of Used Terms

α Chosen neighbor of β

β Current state

b Exponential fit parameter for $r\,(k)$; characteristic length of $r\,(k)$

E_α Energy of state α

$\underline{\underline{\Gamma}}\,(T)$ Matrix of fundamental transition probabilities

$\underline{\underline{\Gamma}}\,(T)^{(m)}$ Matrix of transition probabilities for move order m

h Hamming distance between local minima measured in natural choice steps, i. e in spin flips

L Characteristic system size: 2^L states for spin systems

m Move order

m_c Move order for choosing an $m_e \leq m_c$ uniformly distributed before each move; also denotes the corresponding neighborhood model

m_e Move order for a chain of exactly m natural choices while $\beta \notin N\,(\beta)$

m_i Move order for inert system with $\beta \in N\,(\beta)$ that allows for up to m natural choices per move

m_{opt} Optimal move order for respective model

$N\,(\beta)$ Set of next neighbors of state β for respective neighborhood model

$N^{\,m}\,(\beta)$ Set of m^{th} order neighbors of state β for respective neighborhood model

n Number of annealing steps allowed for optimization

$P(\Delta E, T)$ Metropolis acceptance probability

$\underline{p}^{\,t}$ Probability distribution of states at time t and temperature $T^{\,t}$

$\underline{\underline{\Pi}}^{(\ m)}$ Matrix of probabilities for choosing m^{th} order neighbors in accord with the respective neighborhood definition

$r(k)$ Correlation between initial energy and energy after k (natural) fundamental moves of an annealing run

T Temperature

Glossary

annealing step A temperature update step: A number of moves are made until the system has equilibrated and the temperature is then changed according to the underlying temperature schedule

choice step Choice of a prospective next state from among the neighbors defined by the respective move class model

- **fundamental** Choice of a neighbor from among the fundamental neighbors of a state
- **extended, m^{th} order** Choice of a neighbor from among the m^{th} order neighbors; consists of m successive fundamental choices.
- **natural** Choice of a natural neighbor of a state

move Comprised of a choice step and the application of the acceptance rule to the cost function values of the current and chosen state.

- **fundamental** Move with order $m = 1$: A fundamental choice is made, then the acceptance rule applied.
- **extended, m^{th} order** Move comprised of an extended choice of m^{th} order and the application of the acceptance rule.
- **natural** Fundamental move for order $m_e = 1$

move class Defines the set of neighbors for each state. Also called neighborhood relation.

move class order The (maximum) number m of fundamental choice steps per extended.

neighbor State in the neighborhood of another state.

- **fundamental** Nearest\ first order neighbors of a state within reach by a predefined configuration change;
  For m_e, these are the natural neighbors and for m_i the natural neighbors plus the current state itself
- **m^{th} order** State possibly to be reached by an m^{th} order choice

natural In state space these are the states directly adjacent to the current state in state space, excluding the current state itself.
For spin systems, these are the states differing from the current state by 1 spin flip

neighborhood Set of states that can be reached from the current state by a choice step of the given order.

neighborhood relation Defines the neighborhood for each state. Also called move class.

fundamental Defines the set of fundamental neighbors

m^{th} **order** Defines the set of states possibly to be reached by an m^{th} order choice

natural Defines the set of natural neighbors

Bibliography

[1] S. Toda, R. Kubo and N. Saitô. *Statistical Physics I*, volume 30 of *Solid-State Sciences*. Springer, Heidelberg, 2nd ed. edition, 1992.

[2] D.T. Pham and D. Karaboga. *Intelligent Optimisation Techniques*. Springer, Heidelberg, 2000.

[3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1091, 1953.

[4] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.

[5] S. Abe and Y. Okamoto, eds. *Nonextensive Statistical Mechanics and Its Applications*, volume 560 of *Lecture Notes in Physics*. Springer, Heidelberg, 2001.

[6] T. J. P. Penna. Traveling salesman problem with Tsallis statistics. *Phys. Rev. E*, 51(1):R1–R3, 1995.

[7] B. Andresen, K.H. Hoffmann, K. Mosegaard, J. Nulton, J.M. Pedersen and P. Salamon. On lumped models for thermodynamic properties of simulated annealing problems. *J. Phys. France*, 49:1485–1492, 1988.

[8] A. Franz and K.H. Hoffmann. Optimal annealing schedules for a modified Tsallis statistics. *J. Comput. Phys.*, 176(1):196–204, 2002.

Declaration Of Independent Work

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt, nicht anderweitig zu Prüfungszwecken vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Chemnitz, 26. Juli 2005

Hartwig Ines

Appendix

A Characteristics Of Investigated Spin Systems

In the folowing tables, the first row gives the number of local energy minima. E_{GS} is the ground state energy, E_i are the local minimum energies and h_i the Hamming distances between the respective local minimum and the global one measured in natural choices, i. e. spin flips.

A.1 5 Spins

Systems with 5 spins have $2^5 = 32$ states, each with 5 natural neighbors.

Characteristics	Systems				
	5a	5b	5c	5d	5e
# minima	3	2	2	2	2
E_{GS}	-4.6551	-4.4659	-5.3934	-6.7395	-7.6590
E_1	-3.8299	-2.9024	-3.2766	-1.7964	-5.6590
h_1	3	4	2	2	5
E_2	-2.4046				
h_2	3				

Table A.1: Characteristics for systems with 5 spins

A.2 6 Spins

Systems with 6 spins have $2^6 = 64$ states, each with 6 natural neighbors.

Characteristics	Systems			
	6a	6b	6c	6d
# minima	3	3	2	3
E_{GS}	-7.5056	-7.5157	-7.4260	-6.5592
E_1	-6.2158	-7.1489	-3.4260	-5.9880
h_1	4	3	6	2
E_2	-2.6491	-5.1787		-5.6331
h_2	3	5		2

Table A.2: Characteristics for systems with 6 spins

A.3 7 Spins

Systems with 7 spins possess $2^7 = 128$ states, each of which has 7 natural neighbors.

Characteristics	Systems					
	7a	7b	7c	7d	7e	7f
# minima	3	4	4	4	3	2
E_{GS}	−9.3931	−6.8151	−10.9711	−8.7467	−9.4879	−8.1354
E_1	−8.6269	−6.6694	−6.5059	−8.1382	−3.6525	−6.8002
h_1	3	3	3	2	4	2
E_2	−5.5898	−5.16588	−4.5059	−3.6525		
h_2	3	3	4	4	6	
E_3		−5.0015	0.2611			
h_3		2	6	5		

Characteristics	Systems				
	7g	7h	7i	7k	7l
# minima	2	4	3	3	2
E_{GS}	−9.2830	−9.4049	−9.9167	−8.1799	−10.1099
E_1	−3.2301	−7.9179	−7.2971	−7.2457	−3.6083
h_1	4	2	5	5	4
E_2		−4.8631	−4.5059	−6.1799	
h_2		6	5	7	
E_3		−4.3012	0.2611		
h_3		4			

Table A.3: Characteristics for systems with 7 spins

A.4 8 Spins

Systems with 8 spins possess $2^8 = 256$ states and each of them has 8 natural neighbors.

Characteristics	Systems					
	8a	8b	8c	8d	8e	8f
# minima	4	3	3	3	4	3
E_{GS}	-8.92795	−8.9432	−9.1169	−10.1982	−9.4583	−12.3615
E_1	-7.94092	−8.6688	−8.7353	−8.6036	−8.84619	−12.3050
h_1	4	3	2	5	3	2
E_2	-6.01335	−8.3543	−7.8064	−7.6773	−8.6641	−2.5356
h_2	2	5	2	2	5	6
E_3	-3.23859				−5.4583	
h_3	6				8	

Characteristics	Systems					
	8g	8h	8i	8k	8l	8m
# minima	4	2	4	5	4	6
E_{GS}	−9.3051	−12.3921	−10.7957	−11.5267	−10.3223	−12.9385
E_1	−6.8640	−7.1651	−7.7251	−8.9779	−8.9945	−9.5939
h_1	4	5	3	7	3	3
E_2	−6.3483		−7.4739	−8.6679	−7.6611	−6.8926
h_2	2		6	3	3	3
E_3	−6.2500		−5.7903	−8.3861	−6.9229	−6.7530
h_3	6		7	4	3	4
E_4				−4.3802		−4.9385
h_4				5		8
E_5						−4.4928
h_5						5

Table A.4: Characteristics for systems with 8 spins

A.5 9 Spins

Systems with 9 spins possess $2^9 = 512$ states, each with 9 natural neighbors.

Characteristics	Systems				
	9a	9b	9c	9d	9e
# minima	3	3	3	2	3
E_{GS}	-15.7558	−11.7258	−14.2146	−13.1591	−12.0644
E_1	−7.74411	−10.2429	−13.2410	−12.0635	−11.9835
h_1	7	3	3	6	2
E_2	−4.726217	−10.0637	−8.3333		−6.8417
h_2	8	5	5		5

Characteristics	Systems				
	9f	9g	9h	9i	9k
# minima	7	4	2	2	2
E_{GS}	−10.8982	−12.1952	−14.9165	−13.5373	−13.5497
E_1	−10.4652	−11.9328	−2.9255	−9.1675	−9.9004
h_1	7	2	8	7	3
E_2	−10.0744	−9.2989			
h_2	4	7			
E_3	−9.6864	−9.2905			
h_3	3	3			
E_4	−8.5452				
h_4	6				
E_5	−8.4652				
h_5	2				
E_6	−6.7759				
h_6	6				

Table A.5: Characteristics for systems with 9 spins

B Detailed Results For All Spin Systems

In the following tables, the optimal results, i.e. the results for the optimal move order, are marked by a *.

B.1 5 Spins

Table B.4 gives the results for the different systems and move classes for all 5-spin systems.

System		Move order m				
		1	2	3	4	5
5a	$\overline{E_f}(m_e)$	−4.509	−4.242	−4.652*	−4.242	−4.653
	$\overline{E_f}(m_i)$	−4.487	−4.473	−4.629*	−4.583	−4.628
5b	$\overline{E_f}(m_e)$	−4.336	−4.284	−4.466*	−4.284	−4.464
	$\overline{E_f}(m_i)$	−4.301	−4.458	−4.464*	−4.460	−4.461
	$\overline{E_f}(m_c)$	−4.336	−4.451	−4.463*	−4.464	−4.465
	$\overline{E_f}(m_c)$	−4.509	−4.508	−4.596	−4.568	−4.609*
5c	$\overline{E_f}(m_e)$	−5.364	−5.095	−5.393*	−5.095	−5.393
	$\overline{E_f}(m_i)$	−5.348	−5.390	−5.392*	−5.386	−5.388
	$\overline{E_f}(m_c)$	−5.364	−5.389	−5.392*	−5.392	−5.393
5d	$\overline{E_f}(m_e)$	−6.724	−6.157	−6.739*	−6.158	−6.736
	$\overline{E_f}(m_i)$	−6.707	−6.729	−6.737*	−6.724	−6.727
	$\overline{E_f}(m_c)$	−6.724	−6.734	−6.739*	−6.738	−6.738
5e	$\overline{E_f}(m_e)$	−7.179	−6.650	−7.248	−6.656	−7.643*
	$\overline{E_f}(m_i)$	−7.144	−7.170	−7.206	−7.322	−7.546*
	$\overline{E_f}(m_c)$	−7.179	−7.186	−7.212	−7.332	−7.466*

Table B.1: Mean final energies for 5-spin systems

B.2 6 Spins

Table B.4 gives the results for the different systems and move classes for all 6-spin systems.

System		Move order m				
		1	2	3	4	5
6a	$\overline{E_f}(m_e)$	-7.174	-7.100	-7.442^*	-7.294	-7.445
	$\overline{E_f}(m_i)$	-7.138	-7.163	-7.376	-7.388	-7.416^*
	$\overline{E_f}(m_c)$	-7.174	-7.184	-7.295	-7.321	-7.358^*
6b	$\overline{E_f}(m_e)$	-7.167	-7.315	-7.438	-7.308	-7.461^*
	$\overline{E_f}(m_i)$	-7.131	-7.283	-7.428	-7.405	-7.440^*
	$\overline{E_f}(m_c)$	-7.167	-7.261	-7.359	-7.370	-7.411^*
6c	$\overline{E_f}(m_e)$	-7.328^*	-6.494	-7.294	-6.649	-7.361
	$\overline{E_f}(m_i)$	-7.282^*	-7.078	-7.231	-7.224	-7.286
	$\overline{E_f}(m_c)$	-7.328^*	-7.180	-7.239	-7.280	-7.320
6d	$\overline{E_f}(m_e)$	-6.395	-6.470	-6.522^*	-6.465	-6.522
	$\overline{E_f}(m_i)$	-6.377	-6.550^*	-6.540	-6.540	-6.535
	$\overline{E_f}(m_c)$	-6.395	-6.534^*	-6.531	-6.539	-6.536

Table B.2: Mean final energies for 6-spin systems

B.3 7 Spins

Table B.4 gives the results for the different systems and move classes for all 7-spin systems.

System		Move order m			
		1	3	5	7
7a	$\overline{E_f}(m_e)$	-9.093	-9.305^*	-9.197	-8.754
	$\overline{E_f}(m_i)$	-9.047	-9.250^*	-9.124	
	$\overline{E_f}(m_c)$	-9.093	-9.216^*	-9.208	-9.183
7b	$\overline{E_f}(m_e)$	-6.737	-6.786^*	-6.785	-6.772
	$\overline{E_f}(m_i)$	-6.724	-6.781	-6.779	-6.765
	$\overline{E_f}(m_c)$	-6.737	-6.765	-6.776^*	-6.779
7c	$\overline{E_f}(m_e)$	-10.944^*	-10.922	-10.761	-10.566
	$\overline{E_f}(m_i)$	-10.926^*	-10.920	-10.733	-10.515
	$\overline{E_f}(m_c)$	-10.944^*	-10.956	-10.930	-10.888
7d	$\overline{E_f}(m_e)$	-8.589^*	-8.580	-8.501	-8.431
	$\overline{E_f}(m_i)$	-8.554	-8.628^*	-8.573	-8.504
	$\overline{E_f}(m_c)$	-8.589	-8.651^*	-8.646	-8.630
7e	$\overline{E_f}(m_e)$	-9.379	-9.443^*	-9.327	-9.196
	$\overline{E_f}(m_i)$	-9.379	-9.433^*	-9.338	-9.216
7f	$\overline{E_f}(m_e)$	-8.053^*	-8.051	-7.970	-7.894
7g	$\overline{E_f}(m_e)$	-9.273^*	-9.258	-9.166	-9.023
7h	$\overline{E_f}(m_e)$	-8.998	-9.020	-9.065^*	-8.972
	$\overline{E_f}(m_i)$	-8.911	-8.996	-9.021^*	-8.944
7i	$\overline{E_f}(m_e)$	-9.369	-9.234	-9.425^*	-9.320
	$\overline{E_f}(m_i)$	-9.298^*	-9.178	-9.136	-8.922
7k	$\overline{E_f}(m_e)$	-7.873^*	-7.713	-7.787	-7.861
7l	$\overline{E_f}(m_e)$	-10.054	-10.088^*	-10.030	-9.942
	$\overline{E_f}(m_i)$	-10.021	-10.082^*	-10.038	-9.957

Table B.3: Mean final energies for 7-spin systems

B.4 8 Spins

Table B.4 gives the results for the different systems and move classes for all 8-spin systems.

System		Move order m			
		1	3	5	7
8a	$\overline{E_f}(m_e)$	-8.686^*	-8.673	-8.542	-8.420
	$\overline{E_f}(m_i)$	-8.661^*	-8.660	-8.544	-8.426
	$\overline{E_f}(m_c)$	-8.686	-8.700^*		
8b	$\overline{E_f}(m_e)$	-8.576	-8.613^*	-8.600	
8c	$\overline{E_f}(m_e)$	-8.924	-8.932^*	-8.835	-8.742
8d	$\overline{E_f}(m_e)$	-9.559^*	-9.445	-9.342	-9.188
	$\overline{E_f}(m_i)$	-9.520^*	-9.443	-9.290	-9.108
8e	$\overline{E_f}(m_e)$	-9.053	-9.123^*	-9.112	-9.046
	$\overline{E_f}(m_i)$	-9.027	-9.122^*	-9.060	-8.988
	$\overline{E_f}(m_c)$	-9.053	-9.094^*	-9.103	-9.095
8f	$\overline{E_f}(m_e)$	-12.327^*	-12.306	-12.169	-11.963
8g	$\overline{E_f}(m_e)$	-8.818	-8.913^*	-8.772	-8.607
8h	$\overline{E_f}(m_e)$	-11.916^*	-11.902	-11.670	-11.430
8i	$\overline{E_f}(m_e)$	-10.223	-10.323^*	-10.163	-9.941
8k	$\overline{E_f}(m_e)$	-10.316^*	-10.259	-10.144	-10.022
8l	$\overline{E_f}(m_e)$	-9.806	-9.964^*	-9.669	-9.415
8m	$\overline{E_f}(m_e)$	-12.487^*	-12.438	-12.067	-11.700

Table B.4: Mean final energies for 8-spin systems

B.5 9 Spins

Table B.4 gives the results for the different systems and move classes for all 9-spin systems.

System		Move order m			
		1	3	5	7
9a	$\overline{E_f}(m_e)$	-15.232^*	-15.043	-14.638	-14.167
9b	$\overline{E_f}(m_e)$	-11.333^*	-11.309	-11.117	-10.909
9c	$\overline{E_f}(m_e)$	-13.680	-13.700^*	-13.378	-13.016
9d	$\overline{E_f}(m_e)$	-11.537^*	-11.520	-11.380	-11.2347
9e	$\overline{E_f}(m_e)$	-11.998^*	-11.959	-11.779	-11.536
9f	$\overline{E_f}(m_e)$	-10.343^*	-10.205	-10.019	-9.858
9g	$\overline{E_f}(m_e)$	-12.586^*	12.361	-11.787	-11.369
9h	$\overline{E_f}(m_e)$	-14.864^*	-14.532	-13.711	-13.017
9i	$\overline{E_f}(m_e)$	-12.922^*	-12.727	-12.426	-12.151
9k	$\overline{E_f}(m_e)$	-12.822^*	-12.663	-12.089	-11.593

Table B.5: Mean final energies for 9-spin systems